

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90304 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US01/16450

(22) International Filing Date: 18 May 2001 (18.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/205,515 19 May 2000 (19.05.2000) US

(71) Applicant (for all designated States except US): HUMAN
GENOME SCIENCES, INC. [US/US]; 9410 Key West
Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BIRSE, Charles, E.
[GB/US]; 13822 Saddlevue Drive, North Potomac, MD
20878 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling
Hill Road, Laytonsville, MD 20882 (US).

(74) Agents: WALES, Michele, M. et al.; Human Genome Sci-
ences, Inc., 9410 Key West Avenue, Rockville, MD 20850
(US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished
upon receipt of that report
- with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACIDS, PROTEINS, AND ANTIBODIES

(57) Abstract: The present invention relates to novel proteins. More specifically, isolated nucleic acid molecules are provided encoding novel polypeptides. Novel polypeptides and antibodies that bind to these polypeptides are provided. Also provided are vectors, host cells, and recombinant and synthetic methods for producing human polynucleotides and/or polypeptides, and antibodies. The invention further relates to diagnostic and therapeutic methods useful for diagnosing, treating, preventing and/or prognosing disorders related to these novel polypeptides. The invention further relates to screening methods for identifying agonists and antagonists of polynucleotides and polypeptides of the invention. The present invention further relates to methods and/or compositions for inhibiting or enhancing the production and function of the polypeptides of the present invention.

WO 01/90304 A2

22

Nucleic Acids, Proteins, and Antibodies

This application refers to a "Sequence Listing" that is provided on electronic media in computer readable form pursuant to Administrative Instructions Section 801(a)(i) and as a paper copy. The Sequence Listing forms a part of this description pursuant to Rule 5.2 and Administrative Instructions Sections 801 to 806, and is hereby incorporated in its entirety.

The Sequence Listing is provided as an electronic file (PA131PCTSL.txt, 5,210,863 bytes in size, created on May 18, 2001) on three identical compact discs (CD-R), labeled "COPY 1," "COPY 2," and "CRF." The Sequence Listing complies with Annex C of the Administrative Instructions, and may be viewed, for example, on an IBM-PC machine running the MS-Windows operating system by using the V viewer software, version 2000 (see World Wide Web URL: <http://www.fileviewer.com>).

Field of the Invention

[0001] The present invention relates to novel proteins. More specifically, isolated nucleic acid molecules are provided encoding novel polypeptides. Novel polypeptides and antibodies that bind to these polypeptides are provided. Also provided are vectors, host cells, and recombinant and synthetic methods for producing human polynucleotides and/or polypeptides, and antibodies. The invention further relates to diagnostic and therapeutic methods useful for diagnosing, treating, preventing and/or prognosing disorders related to these novel polypeptides. The invention further relates to screening methods for identifying agonists and antagonists of polynucleotides and polypeptides of the invention. The present invention further relates to methods and/or compositions for inhibiting or enhancing the production and function of the polypeptides of the present invention.

Background of the Invention

[0002] Protein transport is a quintessential process for both prokaryotic and eukaryotic cells. Transport of an individual protein usually occurs via an amino-terminal signal sequence, which directs, or targets, the protein from its ribosomal assembly site to a particular cellular or extracellular location. Transport may involve any combination of several of the following steps: contact with a chaperone, unfolding, interaction with a receptor and/or a pore complex, addition of energy, and refolding. Moreover, an

extracellular protein may be produced as an inactive precursor. Once the precursor has been exported, removal of the signal sequence by a signal peptidase activates the protein.

[0003] Although amino-terminal signal sequences vary substantially, many patterns and overall properties are shared. Recently, hidden Markov models (HMMs), statistical alternatives to FASTA and Smith Waterman algorithms, have been used to find shared patterns, specifically consensus sequences (Pearson, W.R. and D.J. Lipman PNAS 85:2444-48 (1988); Smith, T.F. and M.S. Waterman J. Mol. Biol. 147:195-97 (1981)). Although they were initially developed to examine speech recognition patterns, HMMs have been used in biology to analyze protein and DNA sequences and to model protein structure (Krogh, A. et al. J. Mol. Biol. 235:1501-31 (1994); Collin, M. et al. Protein Sci. 2:305-14 (1993)). HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides and for opening and extending an insertion or deletion. The algorithms are quite flexible in that they incorporate information from newly identified sequences to build even more successful patterns. Other methods exist to identify membrane associated proteins. Klein et al. have developed a method ("ALOM", also called as KKD) to detect potential transmembrane segments in polypeptides (Klein, P. et al. Biochim. Biophys. Acta, 815:468 (1985)). It attempts to identify the most probable transmembrane segment from the average hydrophobicity value over a range of amino acid residues. It predicts whether the segment is a transmembrane segment (INTEGRAL) or not (PERIPHERAL) and thus, can suggest membrane association of a polypeptide.

[0004] Some examples of the protein families which are known to be plasma membrane associated are receptors (nuclear, 4 transmembrane, G protein coupled, and tyrosine kinase), cytokines (chemokines), hormones (growth and differentiation factors), neuropeptides and vasomediators, protein kinases, phosphatases, phospholipases, phosphodiesterases, nucleotide cyclases, matrix molecules (adhesion, cadherin, extracellular matrix molecules, integrin, and selectin), seven transmembrane receptors, ion channels (calcium, chloride, potassium, and sodium), proteases, transporter/pumps (amino acid, protein, sugar, metal and vitamin; calcium, phosphate, potassium, and sodium) and regulatory proteins. Descriptions of some of these proteins (seven transmembrane receptors, kinases, matrix proteins, fibronectins, defensins, EF-hand domain containing

proteins, mac/perforin family members, pancreatic hormones, serine carboxypeptidases, tumor necrosis factors (TNFs)) and diseases associated with their dysfunction follow.

Seven transmembrane receptors-

[0005] The seven transmembrane receptors (also known as heptahelical, serpentine, or G protein-coupled receptors) comprise a superfamily of structurally related molecules. Possible relationships among seven transmembrane receptors (7TM receptors) for which amino acid sequence had previously been reported are reviewed in Probst et al., *DNA and Cell Biology*, 11(1):1-20 (1992). Briefly, the 7TM receptors exhibit detectable amino acid sequence similarity and all appear to share a number of structural characteristics including: an extracellular amino terminus; seven predominantly hydrophobic α -helical domains (of about 20-30 amino acids) which are believed to span the cell membrane and are referred to as transmembrane domains TM 1-7; approximately twenty well-conserved amino acids; and a cytoplasmic carboxy terminus.

[0006] Each 7TM receptor is predicted to associate with a particular G protein at the intracellular surface of the plasma membrane. The binding of the receptor to its ligand is thought to result in activation (i.e., the exchange of GTP for GDP on the α -subunit) of the G protein which in turn stimulates specific intracellular signal-transducing enzymes and channels. Thus, the function of each 7TM receptor is to discriminate its specific ligand from the complex extracellular milieu and then to activate G proteins to produce a specific intracellular signal. Transmembrane domain-3 (TM3) is believed to be essential in signal transduction (Cotecchia et al., *Proc. Natl. Acad. Sci., USA*, 87:2896-2900 (1990)). Other regions may be essential for biological activity as well (Lefkowitz, *Nature*, 265:603-604 (1993)).

[0007] Mutations in the third intracellular loop of one 7TM receptor (the thyrotropin receptor) and in the adjacent sixth transmembrane domain of another 7TM receptor (the luteinizing hormone receptor) have been reported to be the genetic defects responsible for an uncommon form of hyperthyroidism (Parma et al., *Nature*, 365:649-651 (1993) and for familial precocious puberty (Shenker et al., *Nature*, 365:652-654 (1993)), respectively. In both cases the mutations result in constitutive activation of the G protein receptors. Other studies have shown that mutations that prevent the activation of 7TM receptors are responsible for states of hormone resistance which are responsible for diseases such as

congenital nephrogenic diabetes insipidus. See Rosenthal et al., *J. Biol. Chem.*, 268:13030-13033 (1993). Still other studies have shown that several 7TM receptors can function as protooncogenes and be activated by mutational alteration. See, for example, Allen et al., *Proc. Natl. Acad. Sci. USA*, 88:11354-11358 (1991) which suggests that spontaneously occurring mutations in some 7TM receptors may alter the normal function of the receptors and result in uncontrolled cell growth associated with human disease states such as neoplasia and atherosclerosis. Therefore, mutations in 7TM receptors may underlie a number of human pathologies.

Kinases-

[0008] The kinases comprise the largest known group of proteins, a superfamily of enzymes with widely varied functions and specificities. Kinases regulate many different cell proliferation, differentiation, and signaling processes by adding phosphate groups to proteins. Receptor mediated extracellular events trigger the transfer of these high energy phosphate groups and activate intracellular signaling cascades. Activation is roughly analogous to the turning on a molecular switch, and in cases where signalling is uncontrolled, may be associated with or produce inflammation and cancer.

[0009] Almost all kinases contain a similar 250-300 amino acid catalytic domain. The N-terminal domain, which contains subdomains I-IV, generally folds into a two-lobed structure which binds and orients the ATP (or GTP) donor molecule. The larger C terminal lobe, which contains subdomains VIA-XI, binds the protein substrate and carries out the transfer of the gamma phosphate from ATP to the hydroxyl group of a serine, threonine, or tyrosine residue. Subdomain V spans the two lobes.

[0010] The kinases may be categorized into families by the different amino acid sequences (between 5 and 100 residues) located on either side of, or inserted into loops of, the kinase domain. These amino acid sequences allow the regulation of each kinase as it recognizes and interacts with its target protein. The primary structure of the kinase domain is conserved and contains specific residues and identifiable motifs or patterns of amino acids. The serine threonine kinases represent one family which preferentially phosphorylates serine or threonine residues. Many serine threonine kinases, including those from human, rabbit, rat, mouse, and chicken cells and tissues, have been described

(Hardie, G. and Hanks, S. (1995) *The Protein Kinase Facts Books*, Vol 1:7-20 Academic Press, San Diego, CA).

Matrix Proteins-

[0011] The matrix proteins (MPs) provide structural support, cell and tissue identity, and autocrine, paracrine and juxtacrine properties for most eukaryotic cells (McGowan, S.E. (1992) *FASEB J.* 6:2895-2904). MPs include adhesion molecules, integrins and selectins, cadherins, lectins, lipocalins, and extracellular matrix proteins (ECMs). MPs possess many different domains which interact with soluble, extracellular molecules. These domains include collagen-like domains, EGF-like domains, immunoglobulin-like domains, fibronectin-like domains, type A domain of von Willebrand factor (vWFA)-like modules, ankyrin repeat modules, RDG or RDG-like sequences, carbohydrate-binding domains, and calcium-binding domains.

[0012] The diversity, distribution and biochemistry of MPs is indicative of their many, overlapping roles in cell proliferation and cell signaling. MPs function in the formation, growth, remodeling, and maintenance of bone, and in the mediation and regulation of inflammation. Biochemical changes that result from congenital, epigenetic, or infectious diseases affect the expression and balance of MPs. This balance, in turn, affects the activation, proliferation, differentiation, and migration of leukocytes and determines whether the immune response is appropriate or self-destructive (Roman, J. (1996) *Immunol. Res.* 15:163-178).

Fibronectins-

[0013] Fibronectin proteins play a vital role in the structure and function of the extracellular matrix (ECM). Defects in the function of the ECM are thought to be involved in diseases such as osteoporosis, atherosclerosis, arthritis, and fibrotic diseases. Fibronectin enables cells to adhere to the ECM, and influences the growth and migration of cells as well as the organization of the cytoskeleton. As a major component of the ECM, Fibronectin is thought to influence such processes as cellular adhesion and migration, particularly during development, as well as processes such as wound repair (R.O. Hynes, *PNAS*, 96:2588-90 (1999)).

(Hardie, G. and Hanks, S. (1995) *The Protein Kinase Facts Books*, Vol 1:7-20 Academic Press, San Diego, CA).

Matrix Proteins-

[0011] The matrix proteins (MPs) provide structural support, cell and tissue identity, and autocrine, paracrine and juxtacrine properties for most eukaryotic cells (McGowan, S.E. (1992) *FASEB J.* 6:2895-2904). MPs include adhesion molecules, integrins and selectins, cadherins, lectins, lipocalins, and extracellular matrix proteins (ECMs). MPs possess many different domains which interact with soluble, extracellular molecules. These domains include collagen-like domains, EGF-like domains, immunoglobulin-like domains, fibronectin-like domains, type A domain of von Willebrand factor (vWFA)-like modules, ankyrin repeat modules, RDG or RDG-like sequences, carbohydrate-binding domains, and calcium-binding domains.

[0012] The diversity, distribution and biochemistry of MPs is indicative of their many, overlapping roles in cell proliferation and cell signaling. MPs function in the formation, growth, remodeling, and maintenance of bone, and in the mediation and regulation of inflammation. Biochemical changes that result from congenital, epigenetic, or infectious diseases affect the expression and balance of MPs. This balance, in turn, affects the activation, proliferation, differentiation, and migration of leukocytes and determines whether the immune response is appropriate or self-destructive (Roman, J. (1996) *Immunol. Res.* 15:163-178).

Fibronectins-

[0013] Fibronectin proteins play a vital role in the structure and function of the extracellular matrix (ECM). Defects in the function of the ECM are thought to be involved in diseases such as osteoporosis, atherosclerosis, arthritis, and fibrotic diseases. Fibronectin enables cells to adhere to the ECM, and influences the growth and migration of cells as well as the organization of the cytoskeleton. As a major component of the ECM, Fibronectin is thought to influence such processes as cellular adhesion and migration, particularly during development, as well as processes such as wound repair (R.O. Hynes, *PNAS*, 96:2588-90 (1999)).

[0014] Fibronectin is a disulfide-linked dimeric glycoprotein composed of type I, type II, and type III fibronectin repeats. Type I repeats are approximately 45 amino acids in length and are located at the amino- and carboxy-termini of the protein. Type II domains are approximately 40-60 amino acids in length, and contain four conserved cysteines involved in disulfide bonding. It is thought that the type II domains may function in collagen binding. There are approximately 15-17 type III domains, arranged in tandem in the middle of the protein, that are thought to provide elasticity to fibronectin.

Defensins-

[0015] Mammalian defensins are produced by the epidermis and mucosal epithelium as innate effector molecules thought to function in an antimicrobial capacity. Defensins are cytotoxic peptides with a broad range of activity on gram-positive and negative bacteria, fungi, parasites, viruses, and mycobacteria. The two characterized defensins are the alpha and beta defensins. The alpha-defensins are produced by neutrophils and macrophage, while the beta-defensins are produced by epithelia (Singh, P.K., et al., *PNAS*, 95:14961-66 (1998); Lillard, J.W., et al., *PNAS*, 96:651-56 (1999)).

[0016] Defensin peptides range in length from approximately 29 to 35 amino acids, and include six conserved cysteine residues involved in disulfide bond formation and protein folding. The distribution and connection of the cysteine residues differs between the alpha and beta defensins.

EF-hand domain containing proteins-

[0017] Calcium is well known to be essential for cell signaling. However, calcium also plays a role in such cellular processes as protein processing and membrane traffic to and through the Golgi. Many proteins thought to be involved in the binding of calcium accomplish this in part through a protein calcium-binding domain known as the EF-hand domain.

[0018] The domain consists of a twelve residue loop flanked by a twelve residue alpha-helical domain on both sides. In the EF hand loop, the calcium ion is situated in a coordinated pentagonal bipyramidal configuration. An invariant Glutamic acid or Aspartic acid residue provides two oxygens for liganding the calcium ion.

[0019] Proteins containing this domain include aequorin and Renilla luciferin binding protein (LBP), Recoverins, Calmodulin, Calpain small and large chains, Calretinin, Calcyclin, Fimbrin, Serine/Threonine protein phosphatase, and Diacylglycerol kinase, for example.

MAC/Perforin Family Members-

[0020] The Membrane Attack Complex (MAC) is one of the sequentially activated, membrane bound complexes of the complement system used to eliminate diseased or non-compliant cells. Under this system, activated C5b sequentially binds C6 and C7, which insert into cell membranes. This complex then binds one molecule of C8, followed by between 1 and 18 molecules of C9, which polymerizes to generate a transmembrane channel. These transmembrane channels pierce the membrane, increasing the cell's permeability. These channels permit small molecules in the cell to exchange with the medium. Therefore, water is osmotically drawn into the cell, eventually resulting in the cell bursting.

[0021] Similarly, Perforin is a molecule produced by cytotoxic T cells. In the presence of calcium, Perforin polymerizes into transmembrane channels capable of lysing a variety of target cells in a nonspecific manner.

Pancreatic Hormones- Serine Carboxypeptidases-

[0022] Pancreatic hormone (PP) is a peptide of approximately 80 amino acids in length that is generated in pancreatic islets of Langerhans and consequently secreted. Pancreatic hormone is thought to function as a regulator of pancreatic and gastrointestinal functions.

[0023] Representative members of the pancreatic hormones family of proteins include Neuropeptide Y, Peptide YY, and skin peptide YY. These proteins may be useful as therapeutics for controlling secretion of the gonadotropin-releasing hormone, disorders related to feeding, vasoconstrictory actions, and colonic mobility, as well as antibacterial and antifungal activity.

Serine Carboxypeptidases-

[0024] Carboxypeptidases catalyze the hydrolysis of C-terminal residues of polypeptides. Carboxypeptidases are identified either as metallo-carboxypeptidases or serine-carboxypeptidases.

[0025] Serine carboxypeptidases have the ability to hydrolyze peptides as well as peptide amides from the C-terminus, and have a preferential release of a C-terminal arginine or lysine residue. Their subcellular location is usually extracellular or intracellular. The catalytic activity of serine carboxypeptidases is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen bonded to a serine.

Tumor necrosis factors (TNF)-

[0026] Tumor necrosis factors (TNF) alpha and beta are cytokines, which act through TNF receptors to regulate numerous biological processes, including protection against infection and induction of shock and inflammatory disease. The TNF molecules belong to the "TNF-ligand" superfamily, and act together with their receptors or counter-ligands, the "TNF-receptor" superfamily. So far, nine members of the TNF ligand superfamily have been identified and ten members of the TNF-receptor superfamily have been characterized.

[0027] Many members of the TNF-ligand superfamily are expressed by activated T-cells, implying that they are necessary for T-cell interactions with other cell types which underlie cell ontogeny and functions (Meager, A., *supra*).

[0028] Considerable insight into the essential functions of several members of the TNF receptor family has been gained from the identification and creation of mutants that abolish the expression of these proteins. For example, naturally occurring mutations in the FAS antigen and its ligand cause lymphoproliferative disease (Watanabe-Fukunaga, R. *et al.*, *Nature* 356:314 (1992)), perhaps reflecting a failure of programmed cell death. Mutations of the CD40 ligand cause an X-linked immunodeficiency state characterized by high levels of immunoglobulin M and low levels of immunoglobulin G in plasma, indicating faulty T-cell-dependent B-cell activation (Allen, R.C. *et al.*, *Science* 259:990 (1993)). Targeted mutations of the low affinity nerve growth factor receptor cause a disorder characterized by faulty sensory innervation of peripheral structures (Lee, K.F. *et al.*, *Cell* 69:737 (1992)).

[0029] TNF and LT- α are capable of binding to two TNF receptors (the 55- and 75-kd TNF receptors). A large number of biological effects elicited by TNF and LT- α , acting through their receptors, include hemorrhagic necrosis of transplanted tumors, cytotoxicity, a role in endotoxic shock, inflammation, immunoregulation, proliferation and anti-viral responses, as well as protection against the deleterious effects of ionizing radiation. TNF and LT- α are involved in the pathogenesis of a wide range of diseases, including endotoxic shock, cerebral malaria, tumors, autoimmune disease, AIDS and graft-host rejection (Beutler, B. and Von Huffel, C., *Science* 264:667-668 (1994)). Mutations in the p55 Receptor cause increased susceptibility to microbial infection.

[0030] Moreover, an about 80 amino acid domain near the C-terminus of TNFR1 (p55) and Fas was reported as the "death domain," which is responsible for transducing signals for programmed cell death (Tartaglia *et al.*, *Cell* 74:845 (1993)).

[0031] Plasma membrane associated proteins with a predominant tissue expression pattern are important targets for targeted drug delivery, tumor-targeted therapy (e.g., including, but not limited to, radioimmunotherapy) antibody mediated attack of diseased tissues or cancers, and immune mediated cytotoxicity.

[0032] The discovery of new plasma membrane associated proteins and the polynucleotides encoding these molecules thus satisfies a need in the art by not only providing new compositions useful in the diagnosis, treatment, and prevention of diseases associated with cell proliferation and cell signaling, particularly cancer, immune response and neuronal disorders; but also by providing new targets for immune based therapies.

Summary of the Invention

[0033] The present invention relates to novel proteins. More specifically, isolated nucleic acid molecules are provided encoding novel polypeptides. Novel polypeptides and antibodies that bind to these polypeptides are provided. Also provided are vectors, host cells, and recombinant and synthetic methods for producing human polynucleotides and/or polypeptides, and antibodies. The invention further relates to diagnostic and therapeutic methods useful for diagnosing, treating, preventing and/or prognosing disorders related to these novel polypeptides. The invention further relates to screening methods for

identifying agonists and antagonists of polynucleotides and polypeptides of the invention. The present invention further relates to methods and/or compositions for inhibiting or enhancing the production and function of the polypeptides of the present invention.

Detailed Description

Tables

[0034] Table 1 summarizes some of the polynucleotides encompassed by the invention (including cDNA clones related to the sequences (Clone ID NO:Z), contig sequences (contig identifier (Contig ID:)) and contig nucleotide sequence identifier (SEQ ID NO:X)) and further summarizes certain characteristics of these polynucleotides and the polypeptides encoded thereby. The first column provides the gene number in the application for each clone identifier. The second column provides a unique clone identifier, "Clone ID NO:Z", for a cDNA clone related to each contig sequence disclosed in Table 1. The third column provides a unique contig identifier, "Contig ID:" for each of the contig sequences disclosed in Table 1. The fourth column provides the sequence identifier, "SEQ ID NO:X", for each of the contig sequences disclosed in Table 1. The fifth column, "ORF (From-To)", provides the location (i.e., nucleotide position numbers) within the polynucleotide sequence of SEQ ID NO:X that delineate the preferred open reading frame (ORF) that encodes the amino acid sequence shown in the sequence listing and referenced in Table 1 as SEQ ID NO:Y (column 6). Column 7 lists residues comprising predicted epitopes contained in the polypeptides encoded by each of the preferred ORFs (SEQ ID NO:Y). Identification of potential immunogenic regions was performed according to the method of Jameson and Wolf (CABIOS, 4; 181-186 (1988)); specifically, the Genetics Computer Group (GCG) implementation of this algorithm, embodied in the program PEPTIDESTRUCTURE (Wisconsin Package v10.0, Genetics Computer Group (GCG), Madison, Wisc.). This method returns a measure of the probability that a given residue is found on the surface of the protein. Regions where the antigenic index score is greater than 0.9 over at least 6 amino acids are indicated in Table 1 as "Predicted Epitopes". In particular embodiments, polypeptides of the invention comprise, or alternatively consist of, one, two, three, four, five or more of the predicted epitopes described in Table 1. It will be appreciated that depending on the analytical criteria used to predict antigenic determinants, the exact address of the determinant may

vary slightly. Column 8, "Tissue Distribution" shows the expression profile of tissue, cells, and/or cell line libraries which express the polynucleotides of the invention. The first number in column 8 (preceding the colon), represents the tissue/cell source identifier code corresponding to the key provided in Table 4. Expression of these polynucleotides was not observed in the other tissues and/or cell libraries tested. For those identifier codes in which the first two letters are not "AR", the second number in column 8 (following the colon), represents the number of times a sequence corresponding to the reference polynucleotide sequence (e.g., SEQ ID NO:X) was identified in the tissue/cell source. Those tissue/cell source identifier codes in which the first two letters are "AR" designate information generated using DNA array technology. Utilizing this technology, cDNAs were amplified by PCR and then transferred, in duplicate, onto the array. Gene expression was assayed through hybridization of first strand cDNA probes to the DNA array. cDNA probes were generated from total RNA extracted from a variety of different tissues and cell lines. Probe synthesis was performed in the presence of ³³P dCTP, using oligo(dT) to prime reverse transcription. After hybridization, high stringency washing conditions were employed to remove non-specific hybrids from the array. The remaining signal, emanating from each gene target, was measured using a Phosphorimager. Gene expression was reported as Phosphor Stimulating Luminescence (PSL) which reflects the level of phosphor signal generated from the probe hybridized to each of the gene targets represented on the array. A local background signal subtraction was performed before the total signal generated from each array was used to normalize gene expression between the different hybridizations. The value presented after "[array code]:" represents the mean of the duplicate values, following background subtraction and probe normalization. One of skill in the art could routinely use this information to identify normal and/or diseased tissue(s) which show a predominant expression pattern of the corresponding polynucleotide of the invention or to identify polynucleotides which show predominant and/or specific tissue and/or cell expression. Column 9 provides the chromosomal location of polynucleotides corresponding to SEQ ID NO:X. Chromosomal location was determined by finding exact matches to EST and cDNA sequences contained in the NCBI (National Center for Biotechnology Information) UniGene database. Given a presumptive chromosomal location, disease locus association was determined by comparison with the Morbid Map, derived from Online Mendelian Inheritance in Man (Online Mendelian

Inheritance in Man, OMIM™. McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD) 2000. World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim/>. If the putative chromosomal location of the Query overlaps with the chromosomal location of a Morbid Map entry, an OMIM identification number is disclosed in column 10 labeled "OMIM Disease Reference(s)". A key to the OMIM reference identification numbers is provided in Table 5. Column 11 provides the amino acid position of the ALOM hit(s) predicted for the amino acid sequence shown in SEQ ID NO:Y.

[0035] Table 2 summarizes homology and features of some of the polypeptides of the invention. The first column provides a unique clone identifier, "Clone ID NO:Z", corresponding to a cDNA clone disclosed in Table 1. The second column provides the unique contig identifier, "Contig ID:" corresponding to contigs in Table 1 and allowing for correlation with the information in Table 1. The third column provides the sequence identifier, "SEQ ID NO:X", for the contig polynucleotide sequence. The fourth column provides the analysis method by which the homology/identity disclosed in the Table was determined. Comparisons were made between polypeptides encoded by the polynucleotides of the invention and either a non-redundant protein database (herein referred to as "NR"), or a database of protein families (herein referred to as "PFAM") as further described below. The fifth column provides a description of the PFAM/NR hit having a significant match to a polypeptide of the invention. Column six provides the accession number of the PFAM/NR hit disclosed in the fifth column. Column seven, "Score/Percent Identity", provides a quality score or the percent identity, of the hit disclosed in columns five and six. Columns 8 and 9, "NT From" and "NT To" respectively, delineate the polynucleotides in "SEQ ID NO:X" that encode a polypeptide having a significant match to the PFAM/NR database as disclosed in the fifth and sixth columns. In specific embodiments polypeptides of the invention comprise, or alternatively consist of, an amino acid sequence encoded by a polynucleotide in SEQ ID NO:X as delineated in columns 8 and 9, or fragments or variants thereof.

[0036] Table 3 provides polynucleotide sequences that may be disclaimed according to certain embodiments of the invention. The first column provides a unique clone identifier, "Clone ID", for a cDNA clone related to contig sequences disclosed in Table 1. The

second column provides the sequence identifier, "SEQ ID NO:X", for contig sequences disclosed in Table 1. The third column provides the unique contig identifier, "Contig ID:", for contigs disclosed in Table 1. The fourth column provides a unique integer 'a' where 'a' is any integer between 1 and the final nucleotide minus 15 of SEQ ID NO:X, and the fifth column provides a unique integer 'b' where 'b' is any integer between 15 and the final nucleotide of SEQ ID NO:X, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:X, and where b is greater than or equal to a + 14. For each of the polynucleotides shown as SEQ ID NO:X, the uniquely defined integers can be substituted into the general formula of a-b, and used to describe polynucleotides which may be preferably excluded from the invention. In certain embodiments, preferably excluded from the invention are at least one, two, three, four, five, ten, or more of the polynucleotide sequence(s) having the accession number(s) disclosed in the sixth column of this Table. In further embodiments, preferably excluded from the invention are the specific polynucleotide sequence(s) contained in the clones corresponding to at least one, two, three, four, five, ten, or more of the available material having the accession numbers identified in the sixth column of this Table.

[0037] Table 4 provides a key to the tissue/cell source identifier code disclosed in Table 1, column 8. Column 1 provides the tissue/cell source identifier code disclosed in Table 1, Column 8. Columns 2-5 provide a description of the tissue or cell source. Codes corresponding to diseased tissues are indicated in column 6 with the word "disease". The use of the word "disease" in column 6 is non-limiting. The tissue or cell source may be specific (e.g. a neoplasm), or may be disease-associated (e.g., a tissue sample from a normal portion of a diseased organ). Furthermore, tissues and/or cells lacking the "disease" designation may still be derived from sources directly or indirectly involved in a disease state or disorder, and therefore may have a further utility in that disease state or disorder. In numerous cases where the tissue/cell source is a library, column 7 identifies the vector used to generate the library.

[0038] Table 5 provides a key to the OMIM reference identification numbers disclosed in Table 1, column 10. OMIM reference identification numbers (Column 1) were derived from Online Mendelian Inheritance in Man (Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of

Medicine, (Bethesda, MD) 2000. World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim/>). Column 2 provides diseases associated with the cytologic band disclosed in Table 1, column 9, as determined using the Morbid Map database.

[0039]

Definitions

[0040] The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

[0041] In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide. The term "isolated" does not refer to genomic or cDNA libraries, whole cell total or mRNA preparations, genomic DNA preparations (including those separated by electrophoresis and transferred onto blots), sheared whole cell genomic DNA preparations or other compositions where the art demonstrates no distinguishing features of the polynucleotide/sequences of the present invention.

[0042] As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence encoding SEQ ID NO:Y or a fragment or variant thereof; a nucleic acid sequence contained in SEQ ID NO:X (as described in column 3 of Table 1) or the complement thereof; a cDNA sequence contained in Clone ID NO:Z (as described in column 2 of Table 1 and contained within the ATCC Deposit). For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having an amino acid sequence encoded by a

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence contained in Clone ID NO:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide fragment of a polypeptide encoded by SEQ ID NO:X or a polypeptide fragment encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X;

(f) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(g) a polynucleotide which is a variant of SEQ ID NO:X;

(h) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(i) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(j) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(i), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X.

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z;

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X.
4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence contained in cDNA Clone ID NO:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence selected from the group consisting of:
 - (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z;

(e) a full length protein of SEQ ID NO:Y or the encoded sequence contained in cDNA Clone ID NO:Z;

(f) a variant of SEQ ID NO:Y;

(g) an allelic variant of SEQ ID NO:Y; or

(h) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

24. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11.